Weekly Report

9/28/2015-10/04/2015

1 LDA：一般的参数估计方法

我们一般需要解决两个问题[1]:(1)找到一个参数θ,这个参数是能够最好得解释观察变量$\mathcal{X}$.(2)在之前的观测变量下,计算新的观测变量$\tilde{x}$的概率.前面一个被成为参数估计问题,后面一个是预测或者回归问题.观测变量$\mathcal{X}=\{x_i|i=1,2,3...N\}$可以看作一列独立同分布的变量.比如,对于高斯分布来说,$\theta=\{\mu,\sigma^2\}$.

1.1 Maximum likelihood estimation

极大似然估计(Maximum likelihood estimation)是通过最大化似然函数找到参数:

$$L(\theta|\mathcal{X})=p(\mathcal{X}|\theta)=\bigcap_{x\in X}\{X=x|\theta\}=\prod_{x\in X}p(x|\theta)$$

极大似然估计的参数就是使得似然函数最大的参数.

$$\theta_{ML}=\arg\max_{\theta}L(\theta|\mathcal{X})=\arg\max_{\theta}\sum_{x\in\mathcal{X}}\log p(x|\theta)$$

之后就是对θ求导或者用优化算法求参数,具体由函数形式决定.那么,以上解决了问题1,至于问题2,可以由以下逼近:

$$\begin{aligned}
p(\tilde{x}|\mathcal{X}) &= \int_{\theta\in\Theta}p(\tilde{x}|\theta)p(\theta|\mathcal{X})d\theta \\
&\approx \int_{\theta\in\Theta}p(\tilde{x}|\theta_{ML})p(\theta|\mathcal{X})d\theta \\
&= p(\tilde{x}|\theta_{ML})
\end{aligned}$$

我对上式的理解是,实际上应该对所有可能的θ根据它们出现的概率作为权重积分起来,但是这里我们只简单得把最大可能的θ取出来,而不管θ的分布,所以做了近似.

1.2 Maximum a posteriori estimation

Maximum a posteriori estimation被翻译为最大后验估计,这里要用到贝叶斯公式:

$$p(\theta|\mathcal{X}) = \frac{p(\mathcal{X}|\theta)p(\theta)}{p(\mathcal{X})}$$

其中,$p(\theta|\mathcal{X})$,$p(\mathcal{X}|\theta)$,$p(\theta)$,$p(\mathcal{X})$,分别被称为,后验概率,似然函数,先验概率,证据. 我们的目的就是极大化后验概率:

$$\begin{aligned}
\theta_{MAP} &= \arg\max_{\theta} p(\mathcal{X}|\theta) \\
&= \arg\max_{\theta} \frac{p(\mathcal{X}|\theta))p(\theta)}{p(\mathcal{X})} \\
&= \arg\max_{\theta} p(\mathcal{X}|\theta))p(\theta) \\
&= \arg\max_{\theta} \{L(\theta|\mathcal{X}) + \log p(\theta)\} \\
&= \arg\max_{\theta} \{\prod_{x \in X} p(x|\theta) + \log p(\theta)\}
\end{aligned}$$

我们从公式中看到,相比于ML,MAP多了一项先验概率,在实际应用中是为了防止数据被模型过拟合的情况. 所以,对于问题2,类似于上一个方法:

$$\begin{aligned}
p(\tilde{x}|\mathcal{X}) &= \int_{\theta\in\Theta} p(\tilde{x}|\theta)p(\theta|\mathcal{X})d\theta \\
&\approx \int_{\theta\in\Theta} p(\tilde{x}|\theta_{MAP})p(\theta|\mathcal{X})d\theta \\
&= p(\tilde{x}|\theta_{MAP})
\end{aligned}$$

1.3 Bayesian inference

正如之前所说,ML和MAP只是去了一个最大可能的θ,那么是否可以考虑所有可能的θ呢?

$$\begin{aligned}
p(\tilde{x}|\mathcal{X}) &= \int_{\theta\in\Theta} p(\tilde{x}|\theta)p(\theta|\mathcal{X})d\theta \\
&= \int_{\theta\in\Theta} p(\tilde{x}|\theta)\frac{p(\mathcal{X}|\theta)p(\theta)}{p(\mathcal{X})}d\theta
\end{aligned}$$

其中,$p(\mathcal{X}) = \int_{\theta\in\Theta} p(\mathcal{X}|\theta)p(\theta)d\theta$ 贝叶斯参数估计相当于是求一个期望或者说是积分,文章[2]说积分通常是比较难求的,Gibbs采样可以让我们更简单得得到函数分布.这里所写的,也算是对LDA中参数估计以及为什么采用Gibbs的一个介绍.

Plan for next week

- 这周本来想把程序运行起来,但是师兄电脑关了,基站数据库就不能访问了.我想下周应该能把程序跑起来了.

References

[1] G Heinrich. Parameter estimation for text analysis. 2005.

[2] P Resnik and E Hardisty. Gibbs sampling for the uninitiated. 2010.